Collateral Damage of Facebook Apps: Friends, Providers, and Privacy Interdependence

Iraklis Symeonidis, Fatemeh Shirazi, Gergely Biczók, Cristina Pérez-Solà, Bart Preneel

Collateral Damage of Facebook Apps: Friends, Providers, and Privacy Interdependence

Iraklis Symeonidis[1], Fatemeh Shirazi[1], Gergely Biczók[2], Cristina Pérez-Solà[1,3], and Bart Preneel[1]

[1] KU Leuven, ESAT/COSIC and iMinds, Belgium
[2] MTA-BME Future Internet RG, Budapest Univ. of Technology and Economics
[3] Universitat Autònoma de Barcelona, dEIC

Abstract. Third-party apps enable a personalized experience on social networking platforms; however, they give rise to privacy interdependence issues. Apps installed by a user's friends can collect and potentially misuse her personal data inflicting *collateral damage* on the user while leaving her without proper means of control. In this paper, we present a multi-faceted study on the *collateral information collection* of apps in social networks. We conduct a user survey and show that Facebook users are concerned about this issue and the lack of mechanisms to control it. Based on real data, we compute the likelihood of *collateral information collection* affecting users; we show that the probability is significant and depends on both the friendship network and the popularity of the app. We also show its significance by computing the proportion of exposed user attributes including the case of profiling, when several apps are offered by the same provider. Finally, we propose a privacy dashboard concept enabling users to control the *collateral damage*.

1 Introduction

Online Social Networks (OSNs) have become a dominant platform for people to express themselves, interact with each other and get their daily entertainment. By design and popularity, Facebook has morphed into a massive information repository storing users' personal data and logging their interaction with friends, group, events, and pages. The sheer amount and potentially sensitive nature of such data have raised a plethora of privacy issues for Facebook users, such as the lack of user awareness, cumbersome privacy controls, accidental information disclosure, unwanted stalking, and reconstruction of users identities, see Wang et al. [22].

Applications, providers, permissions, and control. Complicating the Facebook privacy landscape, users can also enjoy apps for a personalized social experience. Apps can be developed either by Facebook itself or by third-party app Providers (appPs). Facebook relies on permission-based platform security and applies the least privilege principle to third-party apps. For installation and operation, each app requests from the user a set of *permissions*, granting the app the right to access and collect additional information (steps 1 to 4 in Fig. 1.a). After the user's approval, apps can collect the user's personal data and store it

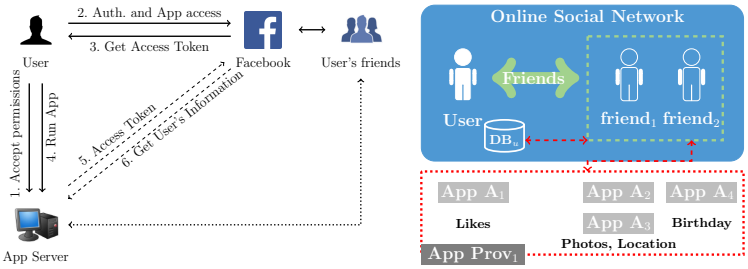

Fig. 1. a. Facebook app architecture, b. *Collateral information collection*

on servers outside Facebook's ecosystem and completely out of the user's control (steps 5 to 6).

Initially, Facebook enabled apps to collect profile attributes of users' friends by assigning separate permissions to each profile attribute. Later, Facebook has replaced this with a single permission to conform with US Federal Trade Commission (FTC) regulations on data collection [3]. Conformity notwithstanding, apps are still able to collect up to fourteen profile attributes via friends [20]. Of course, users have app-related privacy controls at their disposal; however, they are scattered at multiple locations, such as the user's personal profile (visibility levels per attribute) or the apps menu (attributes friends can bring with them to apps). Taking into account that default settings are very much pro-sharing, fragmented and sometimes curiously worded, privacy control settings could promote incorrectly set policies or complete neglect from users [22].

Privacy interdependence, profiling, and legislation. The suboptimal privacy controls and the server-to-server (and potentially offline) communication between Facebook and appP make any protection mechanism hard to apply [9]. As a result, the user's profile items can be arbitrarily retrieved by an appP without automatic notification or on-demand approval by the user through friends. Since the privacy of an individual user is affected by the decisions of other users (being partly out of their control), this phenomenon is referred to as *privacy interdependence* [6]. From an economic point of view, sharing a user's information without her direct consent can lead to the emergence of externalities. While sharing someone else's information may yield benefits for her (positive externality, e.g., personalized experience in social apps), it is also almost certain to cause a decrease in her utility (negative externality, e.g., exposed profile items). Existing research is limited to pointing out the existence of and risks stemming from such negative externalities in the Facebook app ecosystem [6], and its potential impact on app adoption [16, 17].

Neglected by previous work, third party appPs can be owners of several apps (e.g., appP$_1$ offers app A_1, A_2, A_3 and A_4, see Fig. 1.b). For instance, Vipo Komunikacijos and Telaxo are appPs offering 163 and 130 apps, among those 99 and 118 with more than 10,000 monthly active users, respectively (extracted from the Appinspect dataset [5]). As a consequence, an appP may cluster several apps and thus get access to more profile items. Moreover, every app retrieves the Facebook user ID that uniquely identifies a user over apps; hence, the appP could build a combined full profile of the user. We refer to this process as *profiling*, analogously to the term used in the context of consumer behavior in market-

ing [13]. However, with the help of apps installed by a user's friends, appPs could profile a user partly or entirely without her consent, which constitutes a privacy breach, and could induce legal consequences.

From the legal point of view, both the European Data Protection Directive [2] and the guidelines of FTC [3] require prior user consent for the collection and usage of personal data by data controllers (i.e., Facebook or appPs). According to FTC, apps cannot imply indirect consent through privacy settings; while the European Commission requires transparency and fairness from the data controller about the nature, amount, and aim of data collection: this requirement is not met here with data processing potentially going beyond the users' legitimate expectation.

Motivated by the above privacy issues of Facebook apps we define as *collateral damage* the privacy loss inflicted by the acquisition of users' personal data by apps installed by users' friends, and by appPs offering multiple apps thereby enabling user profiling.

Contribution. We have identified four research questions to further our understanding of indirect and *collateral information collection* in the case of Facebook apps.

- *Are the users aware of and concerned about their friends being able to share their personal data?* We conducted an online survey of 114 participants, to identify the users' views on *collateral information collection, lack of notification and not being asked for their approval.* Our survey provides evidence that participants are very concerned and their concern is bidirectional: the large majority of users wants to be notified and potentially restrict apps' access to profile items both when their friends might leak information about them and vice versa.
- *What is the likelihood that an installed app enables the collateral information collection?* We develop a formula to estimate the probability of this event. We show how the likelihood depends on the number of friends and the number of active users of apps. Moreover, based on results obtained from simulations, we show how the likelihood depends on specific network topologies and app adoption models.
- *How significant is the collateral damage*? We develop a mathematical model and quantify the proportion of user attributes collected by apps installed only by the user's friends, including the case of *profiling*, when several apps belong to the same appP. We compute the significance on several snapshots of the most popular Facebook apps using the Appinspect dataset [5].
- *How can we raise user awareness and help them make informed decisions?* For this end, we discuss a dashboard that enhances transparency by providing an overview of installed apps and the type and total amount of profile attributes collected by apps and, more importantly, appPs.

The rest of the paper is organized as follows. Section 2 presents the user survey. Section 3 presents the mathematical model of *collateral damage* and calculates the likelihood of a user being affected by *collateral information collection.* Section 4 extends the model and quantifies *collateral information collection* illustrated by a case study of popular apps. Section 5 presents the high-level design

for a privacy dashboard providing users with proper notifications and control. Section 6 describes future work and concludes the paper.

2 User survey

In this section, we tackle the research question: "are users concerned about *collateral* information collection?" To answer this question, we conducted an online survey investigating users' views about the disclosure of personal data by Facebook apps installed by the users' friends, and to identify users' concerns about unconsented information collection on Facebook; 114 participants answered the survey. Participants were recruited from the authors' direct and extended friend circles (including mostly, but not only, Facebook friends). Hence, a large proportion of participants have an age between 20 and 35 and are well educated. We found that users are concerned about *collateral information collection* in general, and remarkably concerned when information collection is unconsented. Furthermore, the majority of users prefer to take action to prevent *collateral information collection*. We have to stress that our survey provides us with evidence that users are concerned about the information collection of apps through their users' friends. However, we are not able to extrapolate our findings to the general Facebook population.

2.1 Methodology

After a short introduction, our survey consisted of four main parts. First, we assessed users' standpoints and concerns about default privacy settings and the lack of notification for indirect and unconsented information collection. This assessment is necessary to be able to differentiate users who are concerned independent of their intentions to take actions against such practices. The second part of the survey explores what type of personal data on Facebook users find most sensitive. The third part of our survey is twofold: 1) whether users want to be notified when their friends' apps can collect their personal data or when their installed apps can collect personal data of their friends; 2) which actions users prefer to take in such cases. Users replied the survey questions by marking their responses on a scale of 1 to 5 where 1 stands for "not concerned at all and 5 stands for "extremely concerned"; we also provided a text field where necessary. The fourth part of the survey collects demographics and information regarding the participants' use of Facebook apps.

2.2 Results

For the first part, we observe that for all four statements users show concern (see Fig. 2). For example, 66% of users are at least very concerned about the default privacy setting of Facebook that allows apps to collect information from the user's friends. Similarly, 77% of users are at least very concerned about not being notified when their friends enable *collateral information collection* and 67% for not being notified when one of the user's own apps can collect their friends' information. Finally, 81% of users are at least very concerned about *collateral*

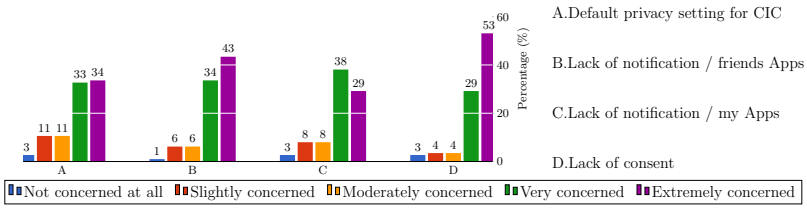

Fig. 2. Results for the first part of the survey where we asked participants about their opinions on four statements regarding default settings, lack of notification (for friends and for the user herself), and lack of consent for *collateral information collection* (CIC).

information collection through apps of their friends without their approval. Note that Golbeck et al. [11] have investigated how informed users are regarding the privacy risks of using Facebook apps. Their findings show that users do not always comprehend what type of data is collected by apps even when they have installed the app themselves. Therefore, it is safe to assume incomplete understanding of apps installed by their friends, which is in line with our results. Note that in Fig. 2, there is a slight difference between participants opinion on statement B on the one hand and statements C and D on the other hand for users which are not concerned. This difference might be because statement B is directly related to the users' information loss. Moreover, statement B would burden the user less than C and D, where action by the users is required.

For the second part of our survey, we found that although users are concerned about a number of attributes, the sensitivity is relatively subjective and differs between users. However, it is noteworthy that certain attributes are standing out and have been marked as sensitive by a large proportion of the participants. For example, most of the users identify photos (84% are at least very concerned), videos (79%), their current location (76%), and family and relationships (54%) as sensitive profile attributes. The least sensitive profile attributes are proved to be to be birthday and sexual orientation. Note that the sensitivity of the attributes is likely to depend on the context. For example, although a birthday attribute might seem harmless on its own, participants might feel different if a weather app would be collecting this information.

In the third part of the survey, we found that 77% of users always want to be notified when friends' apps can collect their personal data, 22% only want to be notified in particular cases, while only about 1% do not want to be notified at all. Moreover, 69% of users always want to be notified when their apps are collecting information from their friends, 27% in particular cases, and only about 1% not at all. We observe that users are also seriously concerned about damaging their friends' privacy: this corroborates findings on other-regarding preferences from the literature [8, 18]. Notification tools can be very useful to enhance privacy awareness for unconsented data collection. Note that Golbeck et al. have shown that the privacy awareness of users can be changed significantly through educational methods [11]. When participants were asked which action they would want to take if notified that friends' apps are about to collect their information (multiple answers allowed), 99 out of 114 participants answered that they would

restrict access to their personal data while 8 participants answered that they would unfriend their Facebook friend. Only 5 participants answered that they would take no action. We have to stress that the reaction of a user may strongly depend on the relationship between the user and their friends. When participants were asked what action they would want to take if they are notified that one of their apps is about to collect their friends' information (multiple answers allowed), 64 out of 114 replied that they would restrict access to their friends' personal information for this app. Only 5 out of 114 answered that they would take no action. The answers to the questions in the third part help to confirm that the answers of our participants in the first part were not due to salience bias; participants who were concerned in the first part about not being notified for the *collateral information collection* replied that they also want to take an action in the third part.

The last part of our survey collected demographics and statistics about Facebook and app usage. Participants were between 16 and 53 years old with an average age of 29 years. They have had their Facebook accounts for between 6 months and 10 years, respectively. Moreover, 69% of our participants have installed an app at least once, and among those 87% have installed 1 or 2 apps in the last six months. 54% of the participants were female, 42% male while 4% preferred not to disclose their gender. Participants varied greatly in their number of friends, from 10 to 1000. 51% changed their privacy settings on Facebook; 79% restricted who could see their profile information, 41% who could see them in searches, and 35% who can collect their information through friends apps (multiple answers were allowed). Interestingly, users who already took an action by restricting their permissions to their friends apps by 90% choose to be notified too. One explanation could be that privacy settings on Facebook are constantly changing and tracking these changes might be cumbersome [22]. Furthermore, 82% of our participants had higher education, where 55% had IT background based on personal interest and 44% through higher education. We conclude from our survey that users are concerned about the *collateral information collection*, and prefer being notified and try to prevent such type of information collection.[4]

3 Likelihood of Collateral Information Collection

In this section, we investigate the likelihood of a user's friend installing an app which enables *collateral information collection*. We build a simple mathematical model and develop a formula to estimate the probability this event occurs. Then, we present case studies taking into account different friendship network topologies and app adoption models. Furthermore, we use the Appinspect dataset [5] to instantiate our estimations, and resort to simulations for computing the probability for different network types.

Let an Online Social Network (OSN) with k users and the corresponding set be denoted by the set $\mathcal{F}$, i.e., $\mathcal{F} = \{u_1, \ldots, u_k\}$. The user is denoted by u, with $u \in \mathcal{F}$. Let f be a friend of u and F^u the set of u's friends, i.e., $f \in \mathsf{F}^u$. Clearly, $\mathsf{F}^u \subseteq \mathcal{F}$. Moreover, let A_j an app and $\mathcal{L}$ the set of all A_js that are offered by the OSN to every u_i, and s the size of the set, i.e., $\mathcal{L} = \{A_1, \ldots, A_s\}$. Moreover, let

[4] `http://http://iraklissymeonidis.info/survey`.

AU_j be the number of users who have installed A_j. For our likelihood estimation we consider the number of Monthly Active Users (MAU) to represent the number of active users. For instance, currently Facebook has $k = 1.3 \times 10^9$ users (i.e., MAU) [19] and more than $s = 25,000$ Apps [5].

To estimate the likelihood that u's personal data can be collected via the A_j, installed by f, we compute the probability of at least an arbitrary f installing any available A_j. Let Q^f be the probability of f installing A_j which enables *collateral information collection*. For all the friends of u (i.e., F^u) the probability of not installing any A_j is the product of probabilities for each f (this assumes that these probabilities are independent, which seems a reasonable approximation). Let Ω be the probability of at least one of u's friends installing A_j (regardless if u has installed A_j), i.e.,

$$\Omega = 1 - \prod_{f \in \mathsf{F}^u} (1 - Q^f) \ . \tag{1}$$

First, we compute the likelihood Ω when the probability for a friend of the user installing an app is uniformly distributed among all friends.

Case study 1 – uniform distribution. Each f decides whether to install A_j without considering any app adoption signals from other users. The probability of at least a friend of u installing A_j is uniformly distributed among u's friends, and equals $1 - Q$ (Remark: $Q = Q^{f_1} = \cdots = Q^{f_{k'}}$ where $1 \leq k' \leq k$). Q is then computed as all users who installed the app divided by the number of users of the OSN (in the active user sense):

$$Q = \frac{AU_j}{|\mathcal{F}|} \ . \tag{2}$$

We used the publicly available Appinspect dataset provided by Hubert et al. [12, 5] to extract the range of MAU of apps which enable *collateral information collection*. The dataset consists of 16,808 Facebook apps between 2012 and 2014. It contains the application name, id, number of active users (daily, weekly and monthly) and the requested permissions. To illustrate the influence of different values of MAUs on Ω, we consider the upper tier of apps, i.e., over 500,000 MAU, while the most popular app that collects friends' data has 10,000,000 MAU, therefore $5 \cdot 10^5 \leq AU_j \leq 1 \cdot 10^7$. To cover most users, we assume the number of friends for a given u ($|\mathsf{F}^u|$) to be between 0 and 1000. Finally, we estimate the population of Facebook to be $1.1 \cdot 10^9$ MAU for the period of 2012 to 2014 [19].

For A_js with $AU_j \geq 5 \cdot 10^6$ the probability Ω grows steeply with the average number of friends (see Fig. 3.a). For an average of 200 friends the probability Ω is more than 0.6. For a user with 300 friends and more, the probability Ω exceeds 0.8. (Note that most Facebook users have more than 200 friends [21].) From Eqns. (1) and (2) it is clear that Ω depends strongly on AU_j. For instance, our most popular app TripAdvisor[5] has approximately $1 \cdot 10^7$ MAU (i.e., $AU_j \approx 1 \cdot 10^7$); assuming that on average a user has 200 friends [21] (i.e., $|\mathsf{F}^u| \approx 200$). Considering $\mathcal{F} = 1.1 \cdot 10^9$ (the population of Facebook) we estimate that the

[5] `https://www.facebook.com/games/tripadvisor`.

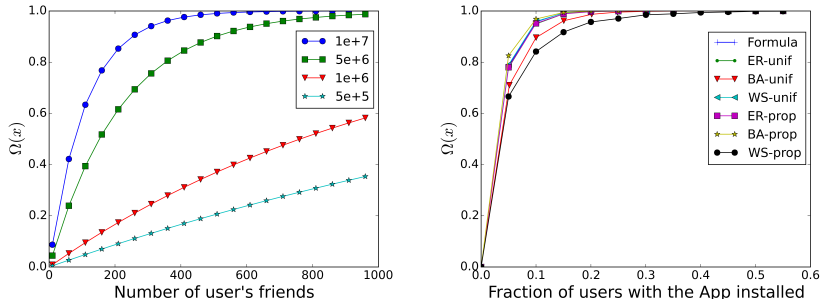

Fig. 3. Likelihood of *collateral information collection* based on a. real data [5] (left, per MAU) and b. simulations (right, with $k = 10,000$ and $d = 30$).

probability of at least one of u's friends installing TripAdvisor is larger than 78% ($\Omega \geq 0.78$).

Case study 2 – non-uniform distribution. Realistic social networks do not conform to the uniformity assumption. Network degree has been reported to follow a power law [15, 24] and the clustering coefficient has been found to be much higher than in random networks [15]. Moreover, app adoption has been proclaimed to be affected by different signals [16]. We have resorted to simulations in order to introduce these factors into the estimation of the probability Ω.

Our simulations generate synthetic networks to compute Ω. Regarding the friendship network, we have considered three different, well-known models: Barabási-Albert [4] (`BA`), Watts-Strogatz [23] (`WS`), and Erdős-Rényi [10] (`ER`). Regarding app adoption, two different models have been implemented: uniform (`unif`), where all users install an app with the same probability (that is, independently of installations by their friends); and preferential (`prop`), where the probability of a user installing an app is proportional to the number of its friends that have already installed the app.

Regarding the simulations, for each of the configurations (pairs of network and app adoption models), we have computed the probability Ω for one of the user's friends installing an app with respect to the fraction of the users of the network that installed the app. To make the results of different network models comparable, we fixed both the number of nodes in the network, k, and the mean degree, d. Then, we tuned the parameters of the models to achieve these properties.

We performed simulations for network sizes $k \in [100,\ 10,000]$ and mean degree $d \in [10,\ 60]$. Due to space constraints we include the results of just one set of simulations, but the conclusions we have drawn can be extrapolated to the other tested settings. Fig. 3.b draws the probabilities obtained from networks with $k = 10,000$ and $d = 30$ (results averaged over 100 simulations) and from the analytical uniform app adoption case. Most of the configurations give probability values very close to those obtained when using the formula; the three exceptions are: `ba-unif`, `ws-prop`, and `ba-prop`. The Barabási-Albert model generates graphs with a few very high degree nodes (hubs) and lots of low degree nodes. When combining networks generated with the Barabási-Albert

model with a uniform app adoption model, the probability for a hub to install an app is the same as for any other node. To the contrary, when combining `BA` with the `prop` app adoption model, hubs have a higher probability of installing the app than non-hubs, since having a higher degree makes them more likely to have (more) friends with the app installed. As a consequence, each installation affects, in mean, more users, and thus Ω increases. Concerning `ws-prop`, the Watts-Strogatz model generates very clustered networks;[6] when an app is installed by a member of a community, it gets adopted by all other members easily. However, each new installation inside the same community implies a small increase on the overall Ω, because most of the users affected by the installation were already affected by installations from other members of the community. We observe that the probability computation (i.e., Ω) is conditioned on both the network and app adoption models. However, we found that there is a significant probability for a user's friend to install an app which enables *collateral information collection* for different networks and app adoption models.

4 Significance of Collateral Information Collection

In this section, we develop a mathematical model and compute the volume of the user's attributes that can be collected by apps and appPs when installed by the users' friends. Our calculations are based on several snapshots of the most popular apps on Facebook using the Appinspect dataset [5].

Users and users' friends. Each user u_i in an OSN (i.e., $u_i \in \mathcal{F}$) has a personal profile where each u can store, update, delete and administer her personal data [7]. A u's profile consists of attributes a_i such as name, email, birthday and hometown. We denote the set of attributes of a u's profile as $\mathcal{T}$ and n as the size of $\mathcal{T}$, i.e., $\mathcal{T} = \{a_1, \ldots, a_n\}$. For instance, Facebook currently operates with a set of $n = 25$ profile attributes. Let F^{u*} be the union of u's friends and the u itself and f^* an element of F^{u*}, i.e., $f^* \in \mathsf{F}^{u*}$. Clearly, $\mathsf{F}^{u*} = \{u\} \cup \mathsf{F}^u$ and $\mathsf{F}^u \cap \{u\} = \emptyset$, as u is not a friend of u. For instance, $\mathsf{F}^{u*} = \{u, f_1, \ldots, f_{k'}\}$ describes a user u and its k' friends, where $1 \leq k' \leq k$.

Applications and Application providers. Let $\mathcal{L}$ be the set of apps an app provider (appP) can offer to every u_i in an OSN and s the size of this set, i.e., $\mathcal{L} = \{A_1, \ldots, A_s\}$. Let A_j, for $1 \leq j \leq s$, be the set of attributes that each A_j can collect, i.e., $\mathsf{A}_j \subseteq \mathcal{T}$. Each A_j is owned and managed by an appP denoted by P_j. The set of A_js that belong to P_j it is denoted by P_j, i.e., $\mathsf{P}_j \subseteq \mathcal{L}$. The set of all P_js is denoted by $\mathcal{AP}$ and m the size of the set, i.e., $\mathcal{AP} = \{P_1, \ldots, P_m\}$. From our analysis we identified $s = 16,808$ apps and $m = 2055$ appPs on Facebook indicating that a P_j can have more than one A_j, i.e., $\mathsf{P}_j = \{A_1 \ldots A_{s\prime}\}$ with $1 \leq s\prime \leq 160$ [5].

4.1 Profiling

Application j. When A_j is activated by f^* (i.e., $f^* \in \mathsf{F}^{u*}$), a set of attributes a_i can be collected from u's profile. We denote by $A_j^{u,\mathsf{F}^{u*}}$ an A_j that users in F^{u*}

[6] The expected clustering coeff. can be adjusted with the rewiring prob. parameter.

installed and as $\mathsf{A}_j^{u,\mathsf{F}^{u*}}$ the set of attributes a_i that $A_j^{u,\mathsf{F}^{u*}}$ can collect from u's profile. Clearly, $\mathsf{A}_j^{u,\mathsf{F}^{u*}} \subseteq \mathsf{A}_j \subseteq \mathcal{T}$. The set of all $A_j^{u,\mathsf{F}^{u*}}$s installed by the users in F^{u*} is denoted by $\mathsf{L}^{u,\mathsf{F}^{u*}}$. Clearly, $\mathsf{L}^{u,\mathsf{F}^{u*}} \subseteq \mathcal{L}$.

We denote by $\vec{a}_i$ a vector of length n which corresponds to a_i, i.e., $\vec{a}_i = [\overset{1}{0}\dots\overset{i}{0}1\dots\overset{n}{0}]$. Moreover, we consider $\vec{A}_j^{u,\mathsf{F}^{u*}}$ as a vector of length n, which corresponds to $A_j^{u,\mathsf{F}^{u*}}$, i.e.,

$$\vec{A}_j^{u,\mathsf{F}^{u*}} = \bigvee_{a \in A_j^{u,\mathsf{F}^{u*}}} \vec{a} \;\Leftrightarrow \vec{A}_j^{u,\mathsf{F}^{u*}}[i] \;= \begin{cases} 1 & \text{if } a_i \;\in A_j^{u,\mathsf{F}^{u*}}, \\ 0 & \text{if } a_i \;\notin A_j^{u,\mathsf{F}^{u*}}, \end{cases} \tag{3}$$

for $1 \leq i \leq n$ and $1 \leq j \leq s$.

Note that:

- $x \cup y = \begin{cases} z = 0 & \text{if } x = y = 0, \\ z = 1 & \text{otherwise.} \end{cases}$ and $\vec{x} \vee \vec{y} = \vec{z}$ where $\vec{x}[i] \vee \vec{y}[i] = \vec{z}[i]$.

For instance, an $\mathsf{A}_j^{u,\mathsf{F}^{u*}} = \{a_1, a_i, a_n\}$ is represented as $\vec{A}_j = \vec{a}_1 \vee \vec{a}_i \vee \vec{a}_n = [\overset{1}{1}0\dots\overset{i}{0}1\dots\overset{n}{0}1]$. It represents the attributes that can be collected by A_j when is installed by f (i.e., the user's friend).

Application provider j**.** Each appP consists of a set of $A_j^{u,\mathsf{F}^{u*}}$s denoted by $\mathsf{P}_j^{u,\mathsf{F}^{u*}}$ which users in F^{u*} installed. Each $\mathsf{P}_j^{u,\mathsf{F}^{u*}}$ can collect attributes of u's profile. To identify which a_is can be collected by P_j we consider $\vec{P}_j^{u,\mathsf{F}^{u*}}$ as a vector of length n (i.e., $n \in \mathcal{T}$), which corresponds to $P_j^{u,\mathsf{F}^{u*}}$, i.e.,

$$\vec{P}_j^{u,\mathsf{F}^{u*}} = \bigvee_{\substack{A \in \mathsf{P}_j^{u,f^*} \\ f^* \in \mathsf{F}^{u*}}} \vec{A}^{u,f^*} = \bigvee_{A \in \mathsf{P}_j^{u,\mathsf{F}^{u*}}} \vec{A}^{u,\mathsf{F}^{u*}} \;. \tag{4}$$

Note that: $\vec{P}_j^{u,\mathsf{F}^{u*}} = \bigvee_{f^* \in \mathsf{F}^{u*}} \vec{P}_j^{u,f^*} = (\vec{P}_j^{u} \vee \vec{P}_j^{u,f_1} \vee \dots \vee \vec{P}_j^{u,f_i})$, where $\mathsf{F}^{u*} = \{u, f_1, \dots, f_i\}$ and $\vec{P}^{u,u} = \vec{P}^{u}$.

The complexity of this operation for all f^* in F^{u*} is $\mathcal{O}(n \times |\mathsf{P}_j^{u,\mathsf{F}^{u*}}|)$.

4.2 Degree of *collateral* information collection

Friends f of u ($f \in F^u$) allow access to u's profile by installing A_js. We denote by $\mathit{\Pi}^u_{A_j^u, A_j^{u,\mathsf{F}^u}}$ the number of attributes that can be collected by A_j exclusively from u's friends (and not through the user herself, i.e., $u \notin \mathsf{F}^u$) . Let $\vec{\mathit{\Pi}}^u_{A_j^u, A_j^{u,\mathsf{F}^u}}$ be a vector of length n which $\mathit{\Pi}^u_{A_j^u, A_j^{u,\mathsf{F}^u}}$ provides, where $n = |\mathcal{T}|$, where

$$\vec{\mathit{\Pi}}^u_{A_j^u, A_j^{u,\mathsf{F}^u}} = \vec{A}_j'^u \bigwedge \vec{A}_j^{u,\mathsf{F}^u} \;. \tag{5}$$

Note that: $\vec{x}' \wedge \vec{x} = [\overset{1}{0}\dots\overset{n}{0}]$ and $\vec{x}' \vee \vec{x} = [\overset{1}{1}\dots\overset{n}{1}]$.

The complexity of this operation for all f^* in F^{u*} is $\mathcal{O}(n^4 \times |\mathsf{A}_j^u| \times |\mathsf{A}_j^{u,\mathsf{F}^u}|)$.

Similarly, we denote by $\vec{\Pi}^u_{P_j^u, P_j^{u,\mathsf{F}^u}}$ the number of attributes that can be collected by P_j exclusively from u's friends in F^u, i.e.,

$$\vec{\Pi}^u_{P_j^u, P_j^{u,\mathsf{F}^u}} = \vec{P}_j'^u \bigwedge \vec{P}_j^{u,\mathsf{F}^u} \quad . \tag{6}$$

4.3 The case of Facebook applications

To examine the problem, we extended our analysis for the apps (i.e., A_js) and appPs (i.e., P_js) on Facebook using the Appinspect dataset [12, 5]. For each A_j, apart from the application name and id, the dataset provides us with the requested permissions and the A_js each P_j owns. We computed the proportion of attributes an A_j and P_j can collect through: 1) the user's friends and the user herself (i.e., *profiling*, F^{u*}) and 2) only the user's friends (i.e., *degree of collateral information collection*, F^u). From 16,808 apps, 1202 enables *collateral information collection*. Our analysis focuses on A_js and P_js that have more than 10,000 MAU; there are 207 and 88 respectively in each category.[7]

Profiling, F^{u*}. Performing the analysis over the dataset, we found that 72.4% of A_js and 62.5% of P_js can collect one attribute from F^{u*}. For all A_js and all P_js, 48.6% and 28.7% of attributes which are considered sensitive by the participants of our survey (such as *photos*, *videos*, *location* and *family-relationships*) can be collected. Considering location related attributes such as *current location*, *hometown*, *work_history* and *education_history*, the proportion of attributes that can be collected are 23.5% from A_js and 23.2% from P_js.

Degree of collateral information collection, F^u. For A_js installed only by F^u, 28.9% of them show a degree of *collateral information collection* equal to 1; similarly, 36.3% of all P_js. Moreover for F^u, we identified that the proportion of sensitive attributes that can be collected from A_js and P_js is 46.8% and 37%, respectively; while the proportion of collectable location related attributes is 22.5% for A_js and 36.9% for P_js.

We conclude that the size of the two sets of sensitive attributes, collected via profiling versus exclusively through friends, are both significant and, surprisingly, comparable to each other. We also found that a considerable amount of attributes concerning the user's location can be collected by either A_js or P_js.

5 Damage Control: Privacy Dashboard

Our survey results have shown that users from our survey are not only concerned about the *collateral information collection*: they also want to be notified and restrict access to their personal data on Facebook. They also consider removing the apps that can cause *collateral damage*. The need for transparency calls for a Transparency Enhancing Technology solution, raising awareness of personal data collection and supporting the users' decision-making on the sharing of personal data [14, 1]. Hence, we propose a dashboard that can help users to manage their

[7] `http://iraklissymeonidis.info/Fb_apps_statistics/`.

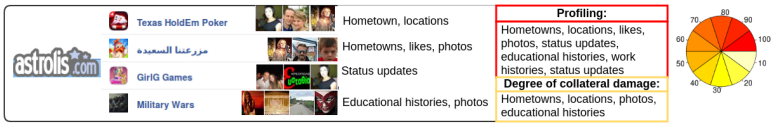

Fig. 4. Privacy dashboard: user interface concept

privacy more efficiently, and control the *collateral information collection* (see Fig. 4 for an initial user interface design). Technically speaking, the dashboard illustrates how the user's data disclosure takes place through the acquisition of the user's personal data via apps (and respective appPs) installed by their Facebook friends. It displays the nature and proportion of the user's personal data that can be collected by apps and, more importantly, appPs.

From our survey, we have concluded that Facebook users are more concerned about certain types of personal information such as photos, videos, location, and relationships. Our dashboard can accommodate the visualization of *profiling* and the degree of *collateral* information collection by assigning different weights to each attribute in the user's profile. These weights can be then manually fine-tuned by the user. Detailed design and implementation of the dashboard remain the next step in our future work. Additional information such as claimed purpose of collection by the apps can be added in the dashboard. Moreover, further functionality can be added to the dashboard such as leading the users from the dashboard to uninstall the app (this would follow the European Data Protection Directive 95/46/EC [2]).

Finally, our survey shows that users also care about the damage that they might cause to their friends by installing apps (bidirectional concern). Complementing the privacy dashboard, we will also look into providing transparency with an enriched app authorization dialogue at the time of installation. Building on the basic design in [22], the enriched dialogue will direct the attention of users to the existence and volume of *collateral damage* to-be-inflicted on their friends.

6 Conclusion and Future Work

In this paper we have presented a multi-faceted study concerning the *collateral damage* caused by friends' apps in social networking sites. Using a user survey, mathematical modeling, and real data from Facebook, we have demonstrated the importance and quantified the likelihood and significance of such *collateral information collection*. Furthermore, to the best of our knowledge, we have been first to report the potential user *profiling* threat that could be achieved by application providers: they can gain access to complementary subsets of user profile attributes by offering multiple apps.

Our main findings are the following. First, our survey shows that the vast majority of users are very concerned and would like proper notification and control mechanisms regarding information collection by apps installed by their friends. Also, they would potentially like to restrict apps' access to profile items both when their friends' apps might collect information about them and vice versa. As for future work, we are aiming at conducting similar surveys among users of social platforms other than Facebook, and extending the demographic range of

participants. We also intend to investigate the relevance of the users concerns and demographic background, attribute values, and sensitivity to particular contexts (e.g., via use cases).

Second, we have quantified the probability that a user is affected by the *collateral information collection* by a friend's app. Assuming a simple app adoption model, an app with more than $500,000$ users may indirectly collect information from the average user with 80% likelihood, irrespective of the user itself having installed the app or not. Moreover, non-uniform app adoption and network models also yield high likelihood. As future work, we aim to extend our simulations regarding both network size and realistic app adoption models.

Third, based on real data, we have quantified the significance of *collateral information collection* by computing the proportion of attributes collected by apps installed by the users' friends. We have found that a significant proportion of sensitive attributes, such as photos, videos, relationships and location, can be collected from apps either by the user's friends and the user herself (i.e., 48.6%) or exclusively from the user's friends (i.e., 46.8%); surprisingly, these values are comparably high. Furthermore, a considerable amount of location-related attributes can be collected by both friends' apps and profiling appPs. As a future work, we aim to enrich our mathematical model by incorporating other parameters such as sensitivity.

Finally, we outline a conceptual design for a privacy dashboard which is able to notify the user about the existence and extent of *collateral damage*, and empower her to take restrictive actions if deemed necessary. We also hint that an enriched app authorization dialogue would complement the dashboard by providing estimates on potential damage to the user's friends at the time of installation. The detailed design and implementation of these solution concepts constitute important future work for us.

Acknowledgments

We notably want to thank Markus Hubert and SBA Research Center for providing us with the necessary material for our study. A thank you to Faruk Gologlu, Filipe Beato, and all the anonymous reviewers who helped for better shaping the idea and the quality of the text. This work was supported in part by the Research Council KU Leuven (C16/15/058), the Spanish Government (TIN2014-55243-P and FPU-AP2010-0078), the Catalan Government (AGAUR 2014SGR-691) and by Microsoft Research through its PhD Scholarship Programme. G. Biczók has been supported by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences.

References

1. Council of the EU Final Compromised Resolution. http://www.europarl.europa.eu. Accessed Feb., 2015.
2. Directive 95/46/EC of the European Parliament and of the Council. http://ec.europa.eu/. Accessed April, 2015.
3. FTC and Facebook agreement for 3rd party apps. http://www.ftc.gov/. Accessed February, 2015.

4. R. Albert and A. Barabási. Statistical mechanics of complex networks. *CoRR*, cond-mat/0106096, 2001.
5. AppInspect. A framework for automated security and privacy analysis of OSN application ecosystems. http://ai.sba-research.org/. Accessed Sept., 2015.
6. G. Biczók and P. H. Chia. Interdependent privacy: Let me share your data. In *17th FC, Okinawa, Japan*, pages 338–353, 2013.
7. D. Boyd and N. B. Ellison. Social network sites: Definition, history, and scholarship. *J. Computer-Mediated Communication*, 13(1):210–230, 2007.
8. D. Cooper and J. H. Kagel. Other regarding preferences: a selective survey of experimental results. *Handbook of experimental economics*, 2009.
9. W. Enck, P. Gilbert, B. Chun, L. P. Cox, J. Jung, P. McDaniel, and A. Sheth. TaintDroid: an information flow tracking system for real-time privacy monitoring on smartphones. *Commun. ACM*, pages 99–106, 2014.
10. P. Erdös and A. Rényi. On the evolution of random graphs. In *Math. Inst. Hungar. Acad. Sci.*, pages 17–61, 1960.
11. J. Golbeck and M. L. Mauriello. User perception of Facebook app data access: A comparison of methods and privacy concerns. *University of Maryland, Maryland*, 2014.
12. M. Huber, M. Mulazzani, S. Schrittwieser, and E. R. Weippl. Appinspect: large-scale evaluation of social networking apps. In *COSN'13, Boston, USA*, pages 143–154, 2013.
13. D. Jobber and F. Ellis-Chadwick. *Principles and practice of marketing*. Number 7th. McGraw-Hill Higher Education, 2012.
14. N. McDonnel, C. Troncoso, P. Tsormpatzoudi, F. Coudert, and L. Métayer. "Deliverable 5.1 : State-of-play: Current practices and solutions." FP7 PRIPARE project. http://priparepproject.eu. Accessed May, 2015.
15. A. Mislove, M. Marcon, P. K. Gummadi, P. Druschel, and B. Bhattacharjee. Measurement and analysis of online social networks. In *7th ACM SIGCOMM, San Diego, USA*, pages 29–42, 2007.
16. Y. Pu and J. Grossklags. An economic model and simulation results of app adoption decisions on networks with interdependent privacy consequences. In *5th GameSec, Los Angeles, CA, USA*, pages 246–265, 2014.
17. Y. Pu and J. Grossklags. Using conjoint analysis to investigate the value of interdependent privacy in social app adoption scenarios. In *36th ICIS*, 2015.
18. D. O. Stahl and E. Haruvy. Other-regarding preferences: Egalitarian warm glow, empathy, and group size. *Journal of Economic Behavior & Organization*, pages 20 – 41, 2006.
19. Statista. Leading Social Networks Worldwide as of January 2016. http://www.statista.com. Accessed Sept., 2015.
20. I. Symeonidis, P. Tsormpatzoudi, and B. Preneel. Collateral damage of Online Social Network Applications. In *2nd ICISSP, Rome, Italy*, 2016.
21. J. Ugander, B. Karrer, L. Backstrom, and C. Marlow. The anatomy of the Facebook social graph. *CoRR*, abs/1111.4503, 2011.
22. N. Wang, H. Xu, and J. Grossklags. Third-party apps on Facebook: Privacy and the illusion of control. In *5th ACM CHIMIT*, pages 4:1–4:10. ACM, 2011.
23. D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):409–10, 1998.
24. C. Wilson, B. Boe, A. Sala, K. P. Puttaswamy, and B. Y. Zhao. User interactions in social networks and their implications. In *4th ACM EuroSys*, pages 205–218, New York, USA, 2009.